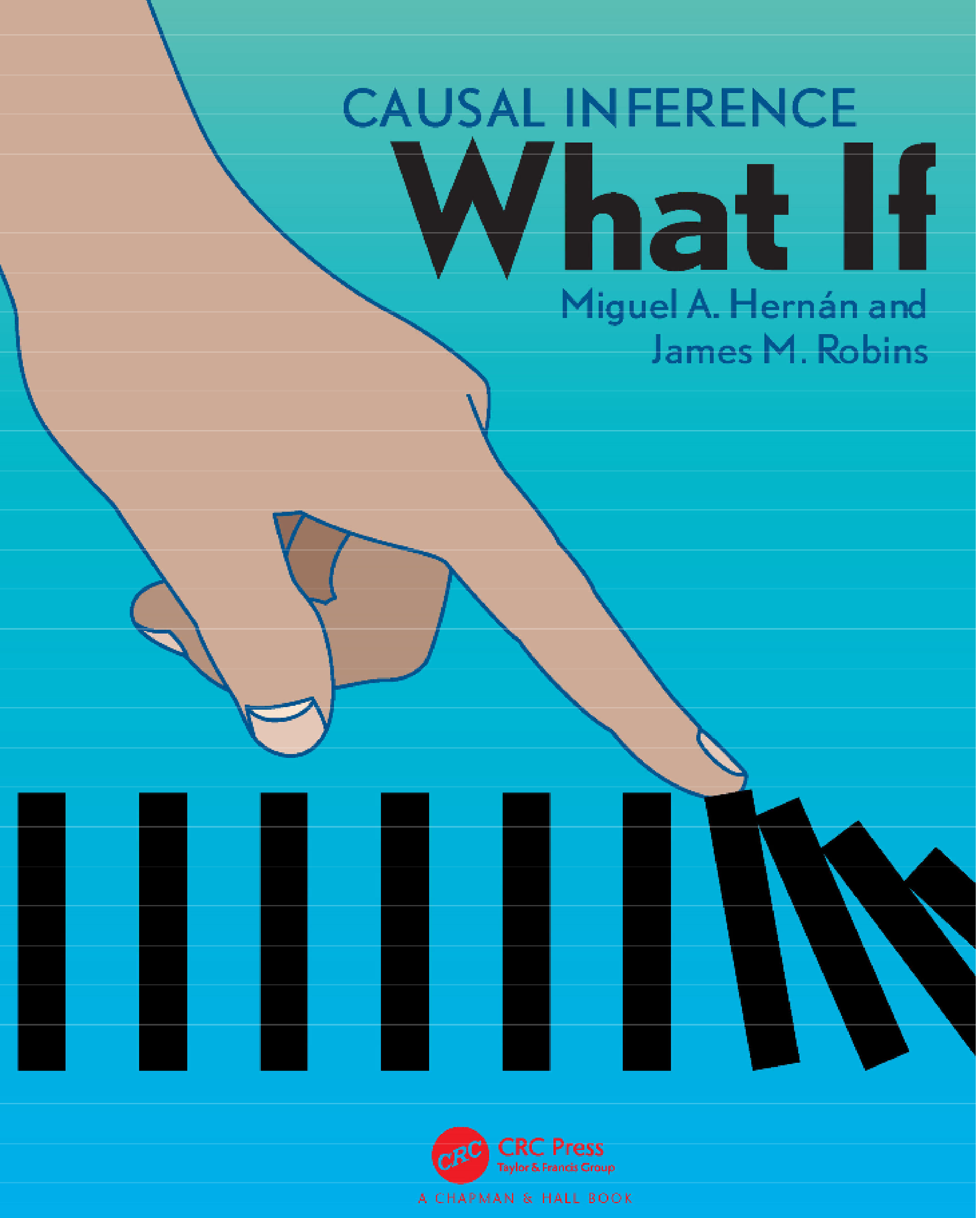



# Chapter 7.1 – 7.2. Confounding

Christina Neuwirth

# Systematic bias

- … when data are insufficient to identify/compute the causal effect even with an infinite sample size
- Informally: any structural association between treatment and outcome that does not come from the causal effect of treatment on outcome

- Confounding: one type of systematic bias
    - common causes – when treatment and outcome share a common cause

# 7.1 The structure of confounding

# Confounding

- Main shortcoming in observational studies
- In observational studies treatment may be determined by many factors

- Structure of confounding can be represented with causal diagrams

# Causal DAG

- L → A → Y


A: Treatment

Y: Outcome

L: Common / shared cause

- 2 sources of associations between A and Y
  - A→ Y: causal effect of A on Y
  - A← L → Y: A and Y are linked through the common cause L

- A← L → Y : **backdoor path**
- Backdoor path: a noncausal path between treatment and outcome that remains even if all arrows pointing from treatment to other variables are removed

# Association is causation

- A→ Y: causal effect of A on Y

 †

- if common cause L did not exist → association between A and Y would be due to the causal effect of A on Y

- Associational risk ratio = causal risk ratio
- $\Pr[Y=1|A=1]/\Pr[Y=1|A=0] = \Pr[Y^{a=1}=1] / \Pr[Y^{a=0}=1]$

# Common cause L

- L → A → Y

- L creates an additonal source of association between A and Y = confounding for the effect of A on Y
- Associational risk ratio != causal risk ratio

- Association is **not** causation

# Examples of the book (I)

**<u>Occupational factors</u>**

- L → A → Y

A: Working as a Firefighter
Y: Risk of death
L: being fit

- → „healthy worker bias“

**<u>Clinical decisions</u>**

- L → A → Y

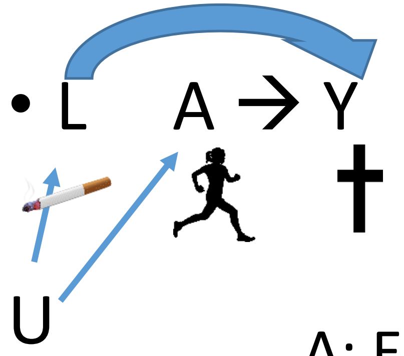
U

A: Aspirin
Y: Stroke
L: heart disease
U: atherosclerosis (unmeasured)

- → „confounding by indication“ or „channeling“

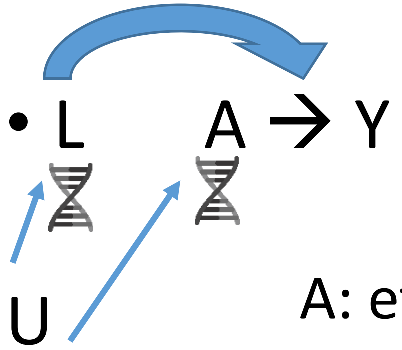
# Examples of the book (II)

## <u>Lifestyle</u>




A: Exercise
Y: Risk of death
L: smoking
U: personality and social factors

## <u>Genetic factors</u>




A: effect of DNA sequence
Y: trait
L: effect of DNA sequence
U: eg ethnicity (linkage of DNA sequences)

- → „linkage desequilibrium“ or „population stratification“

# Examples of the book (III)

**Social factors**

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

A: Income at age 65

Y: Disability at age 75

L: Disability at age 55

**Enviromental exposures**

$$\begin{tikzcd} L \arrow[rr, bend left] & A \arrow[r] & Y \\ U \arrow[u] \arrow[ur] & & \end{tikzcd}$$

A: airborne particular matter

Y: coronary heart disease

L: pollutants

U: weather conditions

# In all cases…

- The bias has the same structure
  - It is due to the presence of a common cause L or U that is shared by A and Y
  - → open backdoor path between A and Y exists

# 7.2. Confounding and exchangeability

# Randomization

- L: individual was in critical condition
- A: heart transplant
- Y: death

- Design 1: randomly select 65% of population and transplant a heart
  - Marginally randomized experiment
- Design 2: divide population into critical and non-critical, then select 75% of population in critical status and 50% in non-critical status
  - Conditionally randomized experiment

# Randomization and exchangeability

- Design 1: Exchangeability means that the risk of death in the first group would have been the same as the risk of death in the second group had individuals in the first group received the treatment given to those in the second group → marginally exchangeable

- Design 2 – critical condition: had the treated remained untreated, their risk of death would have been higher than of those who were actually untreated. → conditionally exchangeable given L

# Exchangeability

- Marginally randomized experiment:
  - Marginal exchangeability holds
  - Average causal effect = calculated as the difference of conditional means $E[Y|A=1]-E[Y|A=0]$
- Conditionally randomized experiment:
  - conditional exchangeabillity holds
    - Average causal effect via adjustment of L via standardization or IP weighting

# If confounding is likely…

- Is there a set of **covariates L** for which conditional exchangeability holds?
- Assumption: we know the true causal DAG
  - Backdoor criterion
  - Transformation of DAG into a Single World Intervention Graph (SWIG)

- Conditional exchangeability holds if **L satifies backdoor criterion:**
  - If **all backdoor paths between A and Y are blocked** by conditoning on L
  - L contains **no variables that are descendants** of treatment A

# Backdoor criterion - example

$$\begin{tikzcd} V \arrow[r] \arrow[d] & W \arrow[d] \\ A \arrow[r] & Y \end{tikzcd}$$

- One backdoor path
- No collider
- Set of variables to control
  - $\{V\}$
  - $\{W\}$
  - $\{V, W\}$

# Backdoor criterion is satisfied when..

- No common cause of treatment and outcome
  - $A \rightarrow Y$
  - Empty set
  - No confounding
  - Marginally randomized experiment where treatment is unconditionally randomized assigned

- Common causes, but a subset L of measured non descendents of A sufficies to block backdoor path
  - $L \rightarrow A \rightarrow Y$
  - Set of variables that satifies backdoor criterion = L
  - Conditionally randomized experiment where probability is the same for all individuals with same value of L

# Backdoor criterion

- Does not measure magnitude or direction
  - Some backdoor paths may be weak and induce little bias
  - Several backdoor paths induce bias in different directions
  - …

- However, it is important to consider the expected direction and magnitude of the bias.

# Thank you!